Psychoneuroendocrinology, Vol. 12, No. 3, pp. 185–192, 1987.
Printed in Great Britain

0306–4530/87 $3.00 + 0.00

REGRESSION TOWARD THE MEAN: MORE ON THE PRICE OF BEER AND THE SALARIES OF PRIESTS

ROBERT D. GIBBONS,*†‡‖ DONALD HEDEKER†§ and JOHN M. DAVIS*†

*Biometric Laboratory, Department of Research, Illinois State Psychiatric Institute, 1601 W. Taylor St., Chicago, IL 60612, U.S.A.; †Department of Psychiatry, University of Illinois, 912 S. Wood St., Chicago, IL 60612, U.S.A.; ‡Department of Biometry, University of Illinois, Chicago, IL 60612, U.S.A.; and §Committee on Research Methodology and Quantitative Psychology, University of Chicago, Chicago, IL 60637, U.S.A.

(*Received* 1 *August* 1986; *in final form* 28 *August* 1986)

SUMMARY

Investigators must take great care in studying the temporal association between biological and clinical data. By selecting patients on the basis of deviant levels of a biological parameter, it often is impossible to separate the effect of regression toward the mean from a temporal association of clinical interest. Experimental strategies that focus exclusively on subjects with extreme values are particularly prone to misinterpreting the effects of regression toward the mean as a temporal change in the variable of interest. We cite specific examples of two previous studies of the relationship between the dexamethasone suppression test (DST) and clinical response in depression. Contrary to the conclusions of the investigators, normalization of the DST had no relation to clinical response in the first 5 weeks of treatment. The decrease of post-dexamethasone cortisol levels occurred regardless of clinical response, most likely due to the effect of regression toward the mean.

IN THE never-ending search to find an application for the dexamethasone suppression test (DST), investigators have turned to an examination of the serial coincidence of DST normalization and clinical response to treatment with antidepressant medication (Greden *et al.*, 1983; Gerken *et al.*, 1985). In a previous commentary (Gibbons & Davis, 1985), we noted that examining mean trend lines for two putatively related measurements tell little or nothing about their actual association. To reiterate, Fig. 1, which is from Greden *et al.* (1983), depicts average trend lines for Hamilton depression rating scale (HAM-D) scores and DST values over the course of treatment. The parallel nature of these two mean trend lines had been taken as evidence for their longitudinal association. Figure 2 represents a similar graph in which parallel mean trend lines are displayed for the annual price of beer and the annual salaries of Protestant ministers from 1973 to 1980. Based on this method of comparison, remarkably similar conclusions must be drawn.

The folly of such an analytic strategy is that the null hypothesis of interest, that is, no association between clinical response and normalization of the DST, is not actually tested by statistically or even visually comparing average trend lines. The only scientifically rigorous test of this hypothesis is to compare DST normalization rates in patients who respond to treatment versus patients who do not. The result of such a comparison in the original data of Greden *et al.* (1983) reveals that DST normalization occurs even more rapidly in patients who did not respond to treatment than in those who did respond

‖Correspondence to be addressed to: Dr R. D. Gibbons, Department of Research, Illinois State Psychiatric Institute, 1601 W. Taylor St., Chicago, IL 60612, U.S.A.

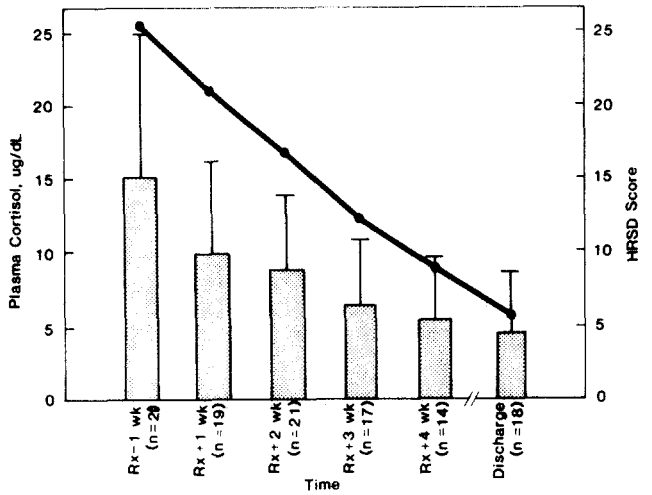

FIG. 1. Mean (± S.D.) maximum post-dexamethasone plasma cortisol concentrations (hatched bars) and corresponding mean Hamilton rating scale for depression (HRSD) scores (connected dots) for dexamethasone non-suppressors. Values are given for last week before treatment (*Rx*), for first four weeks of *Rx*, and for last week of discharge. (From Greden *et al.*, 1983, with permission).

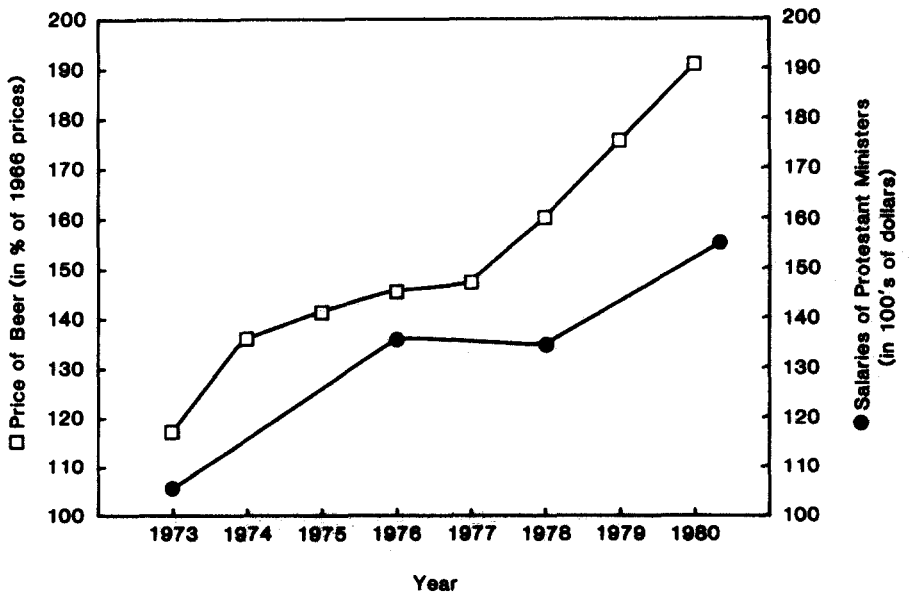

FIG. 2. A plot of the yearly median values of the price of beer and the salaries of Protestant ministers.

(Figure 3).* Hence, time trends may be parallel without indicating that the two variables are causally related. Time trends may be parallel, for example, when the measures are both related to a third variable, such as the overall level of the economy, as in Fig. 2. In this case, once the third variable is covaried out, quite a different relationship between the two variables of interest may be uncovered. Time trends may also be parallel and yet not signify a causal relationship between the two variables when patients are selected on the basis of extreme values, such as non-suppression of cortisol after the administration of dexamethasone. These patients may be expected to have lower values in the future solely due to the effect of "regression toward the mean."

REGRESSION TOWARD THE MEAN

When two variables are perfectly correlated, a unit increase in y will produce a unit increase in x. For example, if parental height x were perfectly correlated with offspring height y, i.e. $r_{xy} = 1.0$, parents whose heights were two standard deviations above the mean height for adults would have a child whose height was two standard deviations above the mean height for children of that age. Alternatively, if there was no association between parental height and offspring height, i.e. $r_{xy} = 0$, the predicted value of the child's height would be the mean value for children his or her age regardless of the heights of the child's parents. Let us, however, make the very reasonable assumption that there exists an association between parental height and offspring height that is neither perfect ($r_{xy} = 1.0$) nor random ($r_{xy} = 0$), but intermediate ($r = 0.5$). If we standardize both parental height and offspring height (i.e., express parental height x and offspring height y both in z score form), we can write the regression of parental height x on offspring height y as:

$$
\begin{aligned}
z_y &= a + bz_x \\
&= (\bar{y} - r_{xy}(s_y/s_x)\bar{x}) + (r_{xy}(s_y/s_x))z_x \\
&= r_{xy}z_x \qquad \text{(since, in } z \text{ score form, the mean } \bar{y} = \bar{x} = 0 \text{ and the standard deviation } s_y = s_x = 1) \\
&= 0.50z_x
\end{aligned}
$$

that is, a child's deviation from the mean height of children that age is one-half as great as his or her parents deviation from the mean height of adults. If for example, parental height is two standard deviations above the mean adult height, the child will have a predicted height of only one standard deviation above the average height of children that age. In other words, tall parents will tend to have children who are above average height, but not as tall as they, and short parents will tend to have short children, but not as short as they. This, of course, is not to say that all parents cannot have children taller than themselves, but that on average they will be shorter. As a historical note, this example of regression toward the mean for offspring height was observed by Sir Francis Galton in the nineteenth century and used in his development of the linear regression model, the term

*The reader should note that the serial association between clinical response and DST normalization was only one of many topics discussed in the report of Greden *et al.* (1983) and that the other topics were investigated with consummate statistical skill. We are further indebted to Dr John Greden for providing us with his data and permitting and encouraging the publication of our findings.

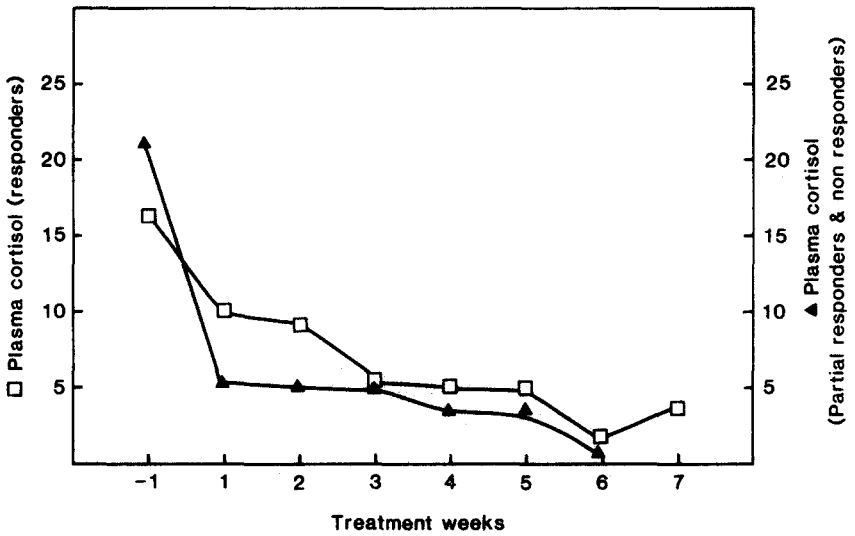

FIG. 3. Actual case: average post-dexamethasone plasma cortisol levels in responders vs non-responders (Greden *et al.*, 1983 data).

"regression" being a shortened form of regression toward the mean in prediction (Hays, 1973, pp. 626–627). Regression toward the mean is characteristic of any relationship in which the correlation is less than perfect. The lower the correlation and the more extreme the value from which the prediction is made, the greater will be the effect of regression toward the mean (Minium, 1970; Cook & Campbell, 1979).

What does all of this have to do with normalization of the DST and its possible relationship to clinical response? If we select patients on the basis of a deviant response at baseline (i.e. high cortisol values following administration of dexamethasone), the predicted value of future cortisol measurements, assuming there is some imperfect relationship between cortisol levels within individuals over time, will tend to be lower independent of any treatment. Consider the data presented by Gerken *et al.* (1985) in Table 1. Even in the first week of treatment, 14 of the 19 patients exhibited decreased cortisol values relative to baseline. It is difficult to imagine that this rapid effect is due to anything but regression toward the mean.

It is difficult to draw any inferences regarding the relationship between DST normalization and clinical response, because all the patients studied by Gerken *et al.* (1985) and Greden *et al.* (1983) recovered or substantially improved with treatment, and all patients also exhibited substantial decreases in their cortisol values over the course of treatment. In order to demonstrate a strong relationship between DST normalization and clinical response, one would need to show that patients who did respond to treatment over time had normalizing cortisol values, as in Fig. 4, and additionally, that patients who did not respond to treatment had non-normalizing cortisol values over time, as in Fig. 5. Thus, without an equivalent sample of patients who did not respond to treatment, i.e. non-responders or patients given placebo, one cannot distinguish between the effects of

TABLE I. HAMILTON RATING SCALE FOR DEPRESSION (HAM-D) AND DST (μg/dl) Gerken *et al.* (1985) DATA

	HAM-D week					DST (μg/dl) week				
Patient	1	2	3	4	5	1	2	3	4	5
1	22	22	17	8	12	7.9	3.7	2.2	1.0	0.9
2	37	26	19	18	10	5.5	2.4	2.0	1.2	0.7
3	28	30	8	7	5	7.2	5.3	6.5	3.3	1.5
4	38	32	37	22	12	6.2	10.9	5.9	3.1	1.7
5	29	25	24	20	15	11.3	17.5	3.2	2.2	3.2
6	27	16	12	18	20	20.1	1.2	0.9	15.4	13.1
7	35	32	22	28	29	10.0	8.6	11.1	11.0	10.1
8	25	29	33	36	34	12.2	18.2	17.3	13.8	8.2
9	19	12	13	7	8	8.4	3.4	2.7	0.0	6.2
10	14	16	6	10	5	13.2	10.6	2.1	2.6	1.0
11	33	45	34	38	28	25.3	16.0	12.7	19.1	10.1
12	37	41	35	34	37	16.8	17.0	20.8	18.9	13.4
13	28	14	17	8	4	10.5	9.6	8.3	7.3	2.0
14	28	23	16	10	8	24.7	12.6	5.2	5.3	6.7
15	30	27	29	31	8	4.8	13.3	6.5	5.2	5.0
16	27	28	13	11	4	16.7	12.4	7.3	17.3	12.8
17	26	22	21	16	4	6.2	2.8	2.2	2.5	0.8
18	19	15	5	5	5	7.3	1.4	0.8	1.3	5.9
19	31	29	31	26	25	17.7	12.1	13.6	10.0	0.9

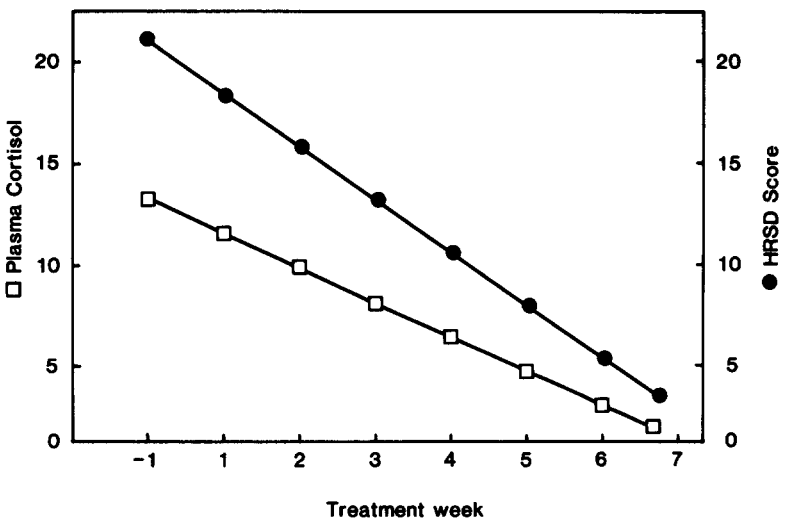

FIG. 4. Ideal case: post-dexamethasone plasma cortisol values and Hamilton rating scale scores in baseline non-suppressors who respond to treatment.

regression toward the mean and an association between cortisol suppression and clinical improvement.

One partial solution to the absence of non-responders is to reanalyze the data of Gerken *et al.* (1985) using only the first 5 weeks of treatment, thereby establishing a group of non-

responders. If we classify responders as patients with a HAM-D score of less than 10 at week 5, there are 10 non-responders and nine responders. Figures 6 and 7 present the mean DST and HAM-D values across the 5 weeks for the responders and non-responders, respectively. The average percent decrease in the DST between week 1 and week 5 was 56.2 ± 32.7% (mean ± S.D.) for the non-responders and 53.0 ± 36.4% for the responders. The difference in means is clearly not statistically significant (t = 0.2, d.f. = 17, p <0.84). In terms of the rates of cortisol normalization over the entire 5 week period, a simple comparison of slopes, or average differences, between the responders and non-responders, computed as:

$$\{\sum_{i=1}^{t-1} (\text{cortisol } wk_{i+1} - \text{cortisol } wk_i)\}/(t - 1)$$

where t = number of weeks, reveals that the slope of the non-responders reflected a decrease of 1.77 ± 1.31 μg/dl/week (mean ± S.D.), whereas the slope of the responders reflected a decrease of 1.59 ± 1.44 μg/dl/week. Again, the results are not significant (t = 0.3, d.f. = 17, p <0.78). As Fig. 8 illustrates, the non-responders exhibited a very similar rate of DST normalization relative to the responders, at least within the first 5 weeks of treatment. Finally a comparison of the DST means at each of the five timepoints by t-test reveals no significant differences between the responders and non-responders.

Thus, the original conclusion of Gerken *et al.* (1985) that DST normalization might precede clinical progress seems doubtful. In fact, when these data are rigorously examined, the conclusion is that DST normalization rates are very similar for both responsive and non-responsive patients. What appears to have been overlooked in the previous analysis was the effect of regression toward the mean in normalizing DST values, regardless of treatment response. Investigators must be particularly aware of the effect of regression toward the mean when samples composed of subjects with extreme values are examined repeatedly.

Supported by a grant from the MacArthur Foundation.

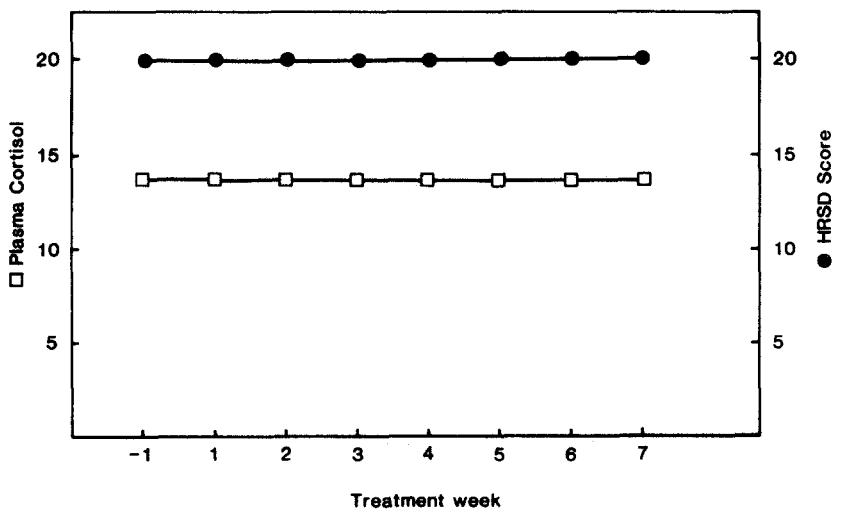

FIG. 5. Ideal case: post-dexamethasone plasma cortisol values and Hamilton rating scale scores in baseline non-suppressors who did not respond to treatment.

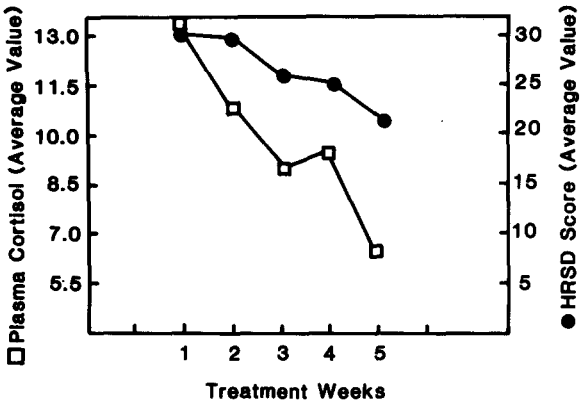

FIG. 6. Actual case: average post-dexamethasone plasma cortisol values and average Hamilton rating scale scores in baseline non-suppressors who did not respond to treatment (n = 10) (Gerken *et al.*, 1985 data).

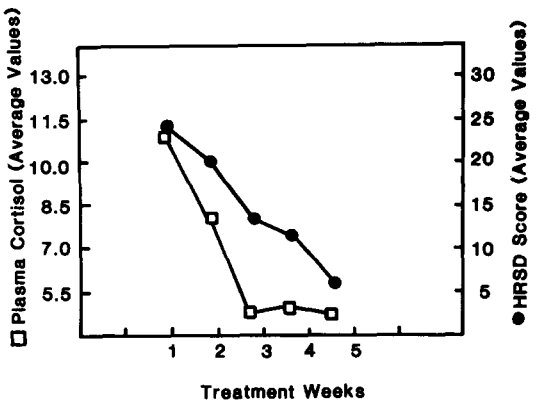

FIG. 7. Actual case: average post-dexamethasone plasma cortisol values and average Hamilton rating scale scores in baseline non-suppressors who responded to treatment (n = 9) (Gerken *et al.*, 1985 data).

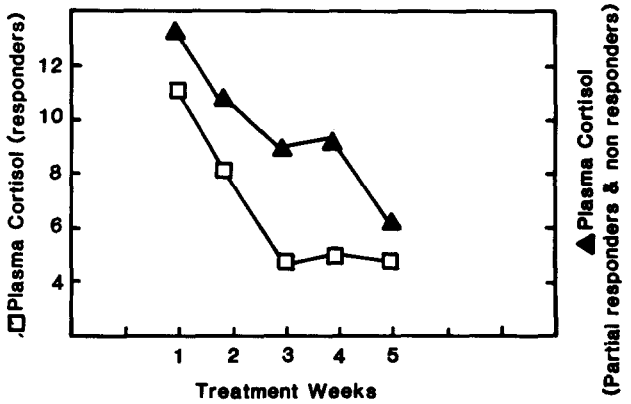

FIG. 8. Actual case: average post-dexamethasone plasma cortisol levels in responders vs non-responders (Gerken *et al.*, 1985 data).

REFERENCES

Cook T D, Campbell D T (1979) *Quasi-Experimentation.* Houghton Mifflin, Boston.

Gerken A, Maier W, Holsboer F (1985) Weekly monitoring of dexamethasone suppression response in depression: its relationship to change of body weight and psychopathology. *Psychoneuroendocrinology* **10:** 261 – 271.

Gibbons R D, Davis J M (1984) The price of beer and the salaries of priests: analysis and display of longitudinal psychiatric data. *Arch Gen Psychiatry* **41:** 1183 – 1184.

Greden J F, Gardner R, King D, Grunhaus L, Carroll B J, Kronfol Z (1983) Dexamethasone suppression test in antidepressant treatment of melancholia: the process of normalization and test-retest reproducibility. *Arch Gen Psychiatry* **40:** 493 – 500.

Hays W L (1973) *Statistics for the Social Sciences,* 2nd edn. Holt, Rinehart, and Winston, New York.

Minium E W (1970) *Statistical Reasoning in Psychology and Education.* Wiley, New York.